DOCUMENT RESUME

ED 132 027 SE 021 556

AUTHOR	Wilson, John T.; And Others
TITLE	Evaluation Report for the Iowa-ASSIST Implementation Effort, 1974-1975.
INSTITUTION	Iowa Univ., Iowa City. Science Education Center.
SPONS AGENCY	National Science Foundation, Washington, D.C.
PUB DATE	Sep 76
GRANT	NSF-GW-8753
NOTE	53p.
EDRS PRICE	MF-$0.83 HC-$3.50 Plus Postage.
DESCRIPTORS	Curriculum; Educational Programs; Educational Research; *Elementary Education; Elementary School Science; Evaluation; Inservice Teacher Education; *Program Evaluation; *Science Education; *Secondary Education; Secondary School Science; *Teacher Workshops
IDENTIFIERS	*Iowa ASSIST

ABSTRACT

This report evaluates the Iowa-ASSIST (Alternative Strategies and Services for Improving Science Teaching) program for assisting schools and communities with improvement of science teaching and science offerings. Program activities that are evaluated include the implementation and revision of major science curriculum programs in Iowa, and inservice teacher workshops designed to provide an introduction to the content, organization, and teaching strategies which characterize a given program. A sample of teacher participants was questioned on the objectives, content, and effectiveness of the workshops. The results of the data analysis led to the conclusion that the workshops were quite effective in fulfilling their objectives. Summative viewpoints on the program are given from both an advocate's view and an adversary's view. (MH)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED132027

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

SE 021 556

project ASSIST

serving education in iowa

Science Education Center, The University of Iowa, Iowa City, Iowa 52242

THE UNIVERSITY OF IOWA

COLLEGE OF EDUCATION

SCIENCE EDUCATION PROGRAM

September, 1976

REPORT NUMBER ONE

Evaluation Report For The Iowa-ASSIST Implementation Effort 1974-1975

by

John T. Wilson
and
Staff

The Iowa-ASSIST Special Report Series

The Iowa-ASSIST Special Report Series was established as a part of the model in-service program from the outset. Program evaluation is an integral part of the program designed to assist with improving science teaching in Iowa. A central staff member has been employed to coordinate all assessment and evaluation efforts. Periodic reports are prepared each year as a record of program achievements. They can be digested and considered over a period of time for major publications.

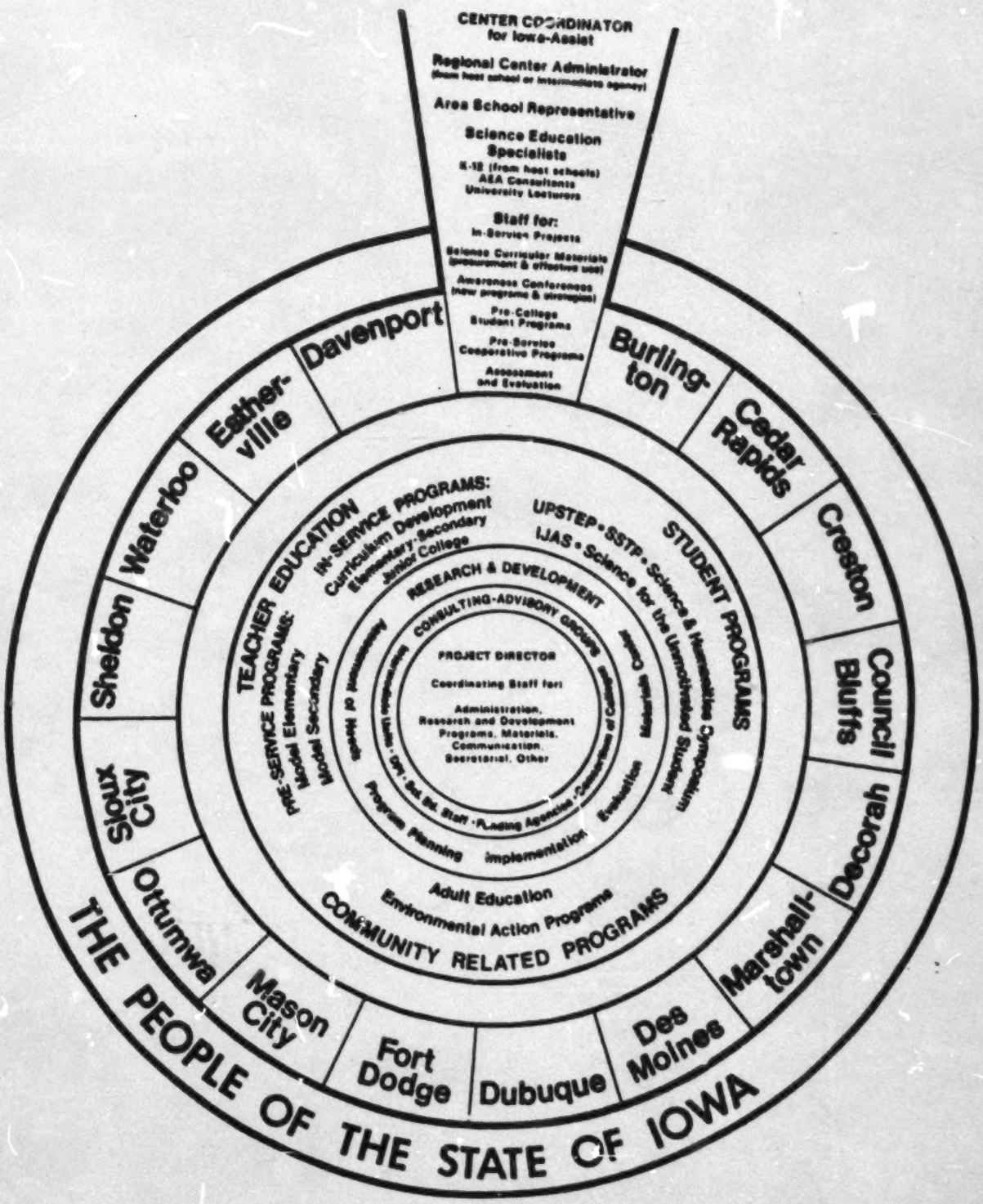
Iowa-ASSIST is a program for assisting schools and communities with improvement of science teaching and science offerings. ASSIST is an acronym for Alternative Strategies and Services for Improving Science Teaching. The program is an integral part of the Science Education Center and represents a major effort with in-service education in Iowa.

Basic to Iowa-ASSIST has been the establishment of sixteen regional service centers throughout the State. A regional coordinator, often a science consultant for the newly created Area Educational Agencies, has been identified and acts in each region to initiate and carry out a variety of improvement efforts. Such coordinators have developed communication networks with all schools (administrators and science teachers) in their geographical areas.

The major activity of Iowa-ASSIST is the assistance provided to schools for directly implementing or revising major science curriculum programs in Iowa schools. Initially the program has provided more assistance with implementing science programs in elementary schools than in secondary schools. Major programs which have been emphasized include ESS, SCIS, SAPA, USMES, MAPS, and Science 5/13. Secondary programs in which assistance has been provided include ISCS, IPS, TPE, IAC, and ISIS. Assistance has been provided in the areas of mathematics and social studies as well. In addition, a K-12 Environmental Studies Program developed by Project ASSIST personnel has proven to be a desirable supplement to existing courses and programs.

Typically the progression of activity starts with an Awareness Conference (usually for administrators and key teachers). Following such conferences, meetings with local teacher groups, community leaders, and central administrators are held as a decision to adopt a given program is made. After such a decision teachers are invited to participate in a special in-service instructional program. Such a program includes a summer introductory workshop (usually two or three weeks in length) which is held at the regional center. Such workshops are designed to provide an introduction to the content, the organization, and the teaching strategies which characterize a given program. Following the summer workshop, a series of in-service meetings (usually six sessions each scheduled for a duration of three hours for each semester) are planned. In addition, school intervisitations, staff visits, and meetings in individual schools all represent informal instructional strategies. The use of the new materials in classrooms with students is central to the academic year phase. Such classrooms represent the "laboratory" for the course(s). Typically each teacher participant earns six semester hours of credit for the summer and the academic year sequence. In some instances the academic year program can be waived if the implementation effort does not require it. The following page represents a pictorial representation of the Iowa-ASSIST program. The organization is complex while flexible; the objectives are general and many-faceted.

Robert E. Yager, Coordinator
Science Education Center
University of Iowa



UNIVERSITY OF IOWA
IOWA-ASSIST
SCIENCE EDUCATION CENTER

John T. Wilson
Assistant Professor
University of Iowa
Coordinator of Instruction and
Evaluation for Iowa-ASSIST

Research Assistants

Masood Ashrafi

Chih-Ru Cheng

Susan Haupt

Ray Reeves

TABLE OF CONTENTS

Rationale 1

Evaluation Instruments and Procedures 3

Instrumentation 4

Data Collection Schedule 6

The Results--Some Preliminary Comments 7

The Nature of the Sample 7

Assessing Participants Content Achievement 8

Content Achievement Objectives Assessed 9

The Design 10

The Instrument 11

Reliability and Validity of Post-Test-I 11

Data Analysis and Results 13

Assessing Affective Outcomes 14

Affective Outcomes Objectives Assessed 15

The Workshop Participant Perceptions 16

Data Analysis and Results 18

Classroom Implementation - Self Reported Data 20

Interpretation 21

Multidimensional Assessment of Philosophy of Education(MAPE). 22

General Description of the MAPE 23

Description of the MAPE Subscales 24

Interpreting MAPE Scores 26

Results 26
Analysis 27
Assessing Participants' Technical Skills 30
Objectives Assessed 30
Pretest-Posttest Assessment 31
Audio-Tape Analysis 33
Self-Report Assessment 35
Summative Viewpoints 37
Advocates View 37
Adversaries View 40
Conclusion 41
References 43

LIST OF TABLES

Table

1 Schedule of Assessment and Evaluation Activities..... 6

2 Descriptive Statistics of the Sample..... 7

3 Technical Data on the Instruments.....12

4 Analysis of Between Group Differences on Posttest Part I.....13

5 Summary of Means, Standard Deviations, and Ranges of Workshop Participant Perceptions.....19

6 Classroom Implementation of Curricular Materials: Summary of Self-Report Type Data.....20

7 MAPE Philosophical Profiles for Seven National Science Curricula.....28

8 Discriminant Scores with Subscales 3, 4, & 5 for Curriculum Populations.....29

9 Pretest-Posttest Analysis for Teaching Skills Inventory --Part III.....32

10 Audio-Tape Analysis for Inquiry Teacher Skills.....34

11 Inquiry Teaching Skills: Summary of Self-Reported Competencies.....36

12 Level of Involvement vs. Teaching Skill Competency.....39

EVALUATION REPORT FOR THE

IOWA-ASSIST IMPLEMENTATION EFFORT

1974-1975

The evaluation reported herein was performed pursuant to grant GW8753 from NSF as part of the Assessment and Evaluation Effort of Project ASSIST/University of Iowa.

EVALUATION REPORT FOR THE
PROJECT ASSIST IMPLEMENTATION EFFORT
1974-1975

Rationale

It was the basic intention of the evaluation effort to determine the extent to which the objectives of the project have been achieved. However, attainment must go beyond a "head count" of participating teachers, administrators and their students and pupils. In the case of Project ASSIST, this number has been impressive. But, if these people were now using the curricula in an inappropriate or inadequate manner, the overall success of the project should be assessed as negative...that is, they would have been better off without the project. This was not the outcome here. This project held as its central objective the acquisition of both knowledge and technical competency by participants relative to teaching a specific science curriculum. In order to assess better this general achievement objective, the following rationale was identified to guide achievement objective, the following rationale was identified to guide the evaluation process:

(1) Evaluation should assess the impact of the project in terms of participant performance. Evaluation data should examine teaching techniques, quality of lessons implemented, responses of learners to lessons implemented, learners' work, and the like. The ultimate measure of success is recognized here to be high quality classroom implementation of curricula.

(2) Evaluation should assess the success of workshop instruction in terms of its objectives. Just as any teacher is concerned with how successful his teaching is in terms of learner achievement and attitude, workshop instructors should also share this concern. Questionnaires and interviews probing why some participating educators do better than others should be used to provide useful feedback to workshop instructors. Close attention should be given to participant perceptions and affective response to workshop instruction since good instruction must at least be attended to by those seeking skills and knowledge.

(3) Evaluation should help teachers and schools assess their impact in terms of the "new" curriculum. Implementation efforts must be evaluated in terms of suitable standards using relevant evaluation procedures. Here these procedures should involve assessing impact in terms of classroom instruction and learner performance. Impact should also assess teacher to teacher influence where teachers using a curriculum help other teachers in the school. How long they continue to use the curriculum is still another dimension of impact.

(4) Evaluation should be objective; criteria for the evaluation process should be available to the public. The greatest implication of "objective evaluation" is openness of the means and processes by which the data were collected and analyzed. How the data are collected should be systematic and objective. Which tests are to be used on the data should be determined in advance. What data are collected

and what conclusions are drawn remain subjective decisions to be guided by the objectives of the project. However, even these subjective decisions should be available for public inspection.

Evaluation in this implementation effort has not been just another instrument or a report tacked on to pre-existing programs for the private concerns of a few. It was instead a process which both supported and was part of many activities in the implementation effort. The data gathered strongly supported the high level of success expected when the project was proposed. The procedures, models, and instruments were not only an effective means for gathering and interpreting data relative to assessing the impact of the effort, but they were also an important instructional tool within the various workshop sessions. They provided important sources of feedback to participants and they monitored participant effort while attempting to utilize a curriculum. Specifically, then, the above rationale describes the approach and materials utilized in the evaluation of the Project ASSIST implementation effort.

Evaluation Instruments and Procedures

A systematic approach to training and evaluation permitted the application of similar techniques and procedures to all target curriculum areas with only minor changes in evaluation and instructional procedures. Instructional teams from each curriculum area were involved in selecting and identifying key aspects of each curriculum program. Often instructional staff were trained to utilize new and

novel techniques in order that relevant and meaningful data be obtained. Many evaluation procedures and instruments were later modified and developed further reflecting the type of workshop activity being carried out which were congruent with the project objectives. This mutual interaction between instruction and evaluation staffs was basic to maintaining mutually relevant evaluation and instructional activities.

Producing, scoring, revising, and maintaining a supply of instructional and evaluation materials were major responsibilities of the Director of Assessment and Evaluation and his research assistants. As the data were collected, periodic summaries and findings were provided, offering suggestions and feedback to instructional teams. Subsequent material revision, instrument development, and instructional material production relevant to techniques of teaching were made as needed. Hence, all of the schedules, materials, and instruments cited and summarized in this report were the result of cooperative efforts of all persons associated with this implementation effort.

Instrumentation

The main function of the evaluation instruments was to sample widely participant knowledge, performance, and attitudes relative to the goals of the project. Three types of instruments were devised which would provide data from all participants. These types of instruments included the following:

(1) Written measures were used to assess the acquisition of factual and conceptual knowledge relative to specified outcomes. Where suitable and available, standardized tests from the curriculum project itself were utilized as a basis to assess the participant's acquisition of curriculum content. The Post-test was the primary written measure utilized.

(2) Performance measures provided a means to monitor and assess the level of teacher participation and performance in a variety of activities. Generally, these measures relied upon the commonality among the teaching strategies utilized in the various curricula. This commonality was identified as the teaching strategy known as "inquiry" and the teaching skills associated with and basic to this strategy, known as "questionning behavior". Protocols and rating procedures were utlized to interpret and assess participant performance relative to the attainment of objectives which implied the acquisition of new behaviors. Audio-taping the teaching of target curricula and reports of teaching activities were the primary source of data describing performance. Although classroom visits were performed, these were generally supplementary and informal.

(3) Measures of Affect included both a sampling of participant attitudes toward the workshops and a measure of the participant's educational philosophy. A procedure known as the "Workshop Participant's Perceptions Survey" was developed which monitored participant's perceptions of workshop activity. The "Multidimensional Assessment of Philosophy of Education (MAPE)" was used to assess

educational philosophies. It is expected that the findings from use of the MAPE will be important in interpreting long-range impact.

Data Collection Schedule

Evaluation of the project required that data be gathered in view of and as part of the instructional transactions planned to attain specific objectives. The schedule as summarized in Table 1 below outlines where and when various data-producing activities occurred. Much of the data was gathered by the instructors. However, formal

TABLE 1

Schedule of Assessment and Evaluation Activities

Procedures	Summer	Academic Year
Written Measures		Post-Test
Performance Measures	Introductory Audio-Taping Sessions	Four or Six Teaching Sessions - Audio-Taped and Analyzed Teaching Activities Reports-Minimum of Three (3) Each Semester
Affective Measures	Workshop Participation Perceptions Weekly	Workshop Participation Perception-Twice Each Semester M.A.P.E.

Assessment was managed by the evaluation staff of Project ASSIST. Scoring, analyzing, and interpreting were accomplished by the evaluation staff of Project ASSIST.

The Results--Some Preliminary Comments

The volumes of data gathered in this project warrant a few comments explaining how the data will be presented. Basically, specific objectives of the project can be classified into three major components: content, attitudes, and technical skills. The data will be presented reflecting these divisions. Some workshops were held in summer only and some were carried on a short term basis. Several participants attended only one or another segment of a workshop. Although we have collected and analyzed data concerning such workshops and participants, these data are not included in this report.

The Nature of the Sample

Most teachers in the workshops as reported in Table 2 below hold a bachelors degree and have taught science for more than six (6) years. Only 7% of ESS participants had a masters degree while in

TABLE 2

Descriptive Statistics of the Sample*

	Degree			Years Exp.			Level Taught			Reason Attended**		
Curricula	Less	BA	MA	0-2	3-6	6-9	k-3	4-6	7-12	A	B	C
SAPA	0	76	19	13	16	65	54	33	0	32	12	22
SCIS	11	69	17	9	22	68	58	37	1	39	41	8
ESS	8	80	7	15	39	43	59	33	2	38	33	7
ISCS	0	65	29	30	12	53	0	6	88	59	10	0
ES	7	66	23	28	30	38	27	30	29	27	30	29

*Percent signs omitted.
**A=Improve curriculum in use, B=Implement new curriculum, C=Earn Credit.

other workshops, i.e., SAPA, ISCS, SCIS, and ES, this percentage was close to 20%. There were some who had not gone beyond high school. Almost all teachers in ESS, SCIS, and SAPA were teaching in K-6 grade levels but ES teachers taught anywhere from K through 12, in ISCS, 59% taught in 6-9 levels and 29% taught in 9-12 levels.

The main reason teachers attended workshops depended on which workshop they attended. Some simply needed the credit. Many teachers however, reported that they wanted to improve the curriculum they were using and many others were interested in implementing new curricula.

Assessing Participants Content Achievement

Evaluation often asks "To what degree do curriculum workshops fulfill the objectives claimed by the instructor". However, a more basic question is "Do workshops have any effect at all?" "Do teachers who attend curriculum workshops know more about the curriculum than those who don't?". It is not unusual for an interested teacher to look around, study, think, and experiment in order to improve his teaching skills and in order to improve the quality of the lessons he is conducting. Such a teacher is very likely to encounter a curriculum in his field, read about it, observe it, even try it in his classes. It is also possible for a teacher to do all these without even hearing about a workshop. Then why do we need any curriculum workshops?

In order to answer this question a test was constructed on some of the objectives of the workshops and given to the participants

during the Spring of 1975, to see how well the objectives were achieved. The same test was also given to teachers who were not attending that workshop. The results of these two experiments were compared to see if the workshops made a significant effect on participant achievement.

Content Achievement Objectives Assessed

The participant, after completing the workshop, should have a general knowledge about the history, the development, the materials, the equipment, the instructional strategies, and the content of the curriculum. In terms of behaviors, the participant should be able to:

a) recognize the materials and equipment related to the curriculum and methods of using them.

b) identify the instructional strategies and techniques recommended by the authors.

c) identify the evaluation procedures and techniques recommended by the authors.

d) identify specific factual content, general conceptual content and procedures, and distinguish between these content areas.

e) (when presented with unfamiliar examples) identify the concept represented.

The participant, after completing the workshop, has general understanding of the philosophy and the development of the curriculum. In terms of the behaviors, he should be able to:

a) recognize those features of the curriculum which are similar to those of other programs.

b) distinguish those features of the curriculum which are not common to other similar programs.

c) select the appropriate implication of a part of the philosophy from other inappropriate ones.

d) identify persons (or groups of persons), who were involved in the development of the topics; recognize firms or companies involved in preparing the materials; and distinguish valid reasons for the development of these materials from invalid ones.

The Design

A posttest-only-control-group design was utilized. The posttest instrument was mailed in May, 1975, to all participants who completed both the summer and academic year phase of the various workshops held across the State of Iowa. A total of 413 out of 560 of the tests were returned in a mailer provided. Some 20 packets were returned for wrong or change of address. A few were lost in the mailing process and some teachers had already left for summer vacations. Considering all these factors a 74% return after a first mailing is quite adequate. The same instrument was sent to a randomly selected group of teachers who had not taken that workshop. The control groups were also participants of the Project ASSIST Workshops. For example an ESS Post-Test-I was sent to all the participants of ESS workshops. This group served as the experimental group. The ESS Post-Test-I was also sent to a number of individuals who were attending other workshops like SAPA or ISCS, who served as the control group. The control group was randomly selected from participants of all other workshops. This

design was chosen because:

(a) it was the most practical one at the time,

(b) it did not require a pre-test and hence the members were not influenced by an advance notice,

(c) the interpretation of the results were simple,

(d) the population sampled for the control group held professional characteristics to those of the experimental group.

The Instrument

For each workshop, an objective test of 30 multiple-choice and/or true-false items was constructed by an evaluator external to the workshop. The Assessment and Evaluation staff of Project ASSIST was responsible for building these instruments. Although the main purpose of the Post-Test-I was to measure the achievement of the content acquisition objectives, the staff decided to include in the test one or two objectives from the technical training section. All the tests were based on the Objectives For Curriculum Workshops.

Reliability and Validity of Post-Test-I

The reliability coefficients of the test, as reported in Table 3 on the next page, are generally low. But one must keep in mind that the reliability of a test is mainly a function of the performance of the group. A low reliability coefficient in this case means that the test is not reliable for discriminating among the individuals attending a particular workshop. Since the tests were

TABLE 3

Technical Data on the Instruments

Workshop	r	Standard Error	Item Diff.	Item Discr.	Range
Experimental					
SAPA	.64	1.91	.62	.25	25 - 12
ESS	.37	1.99	.67	.21	26 - 11
ISCS	.27	1.59	.39	.15	15 - 8
SCIS	.49	2.13	.57	.24	25 - 10
ES	.58	2.10	.38	.24	22 - 8
Control					
SAPA	.83	2.08	.49	.39	21 - 00
ESS	.81	2.11	.58	.40	24 - 0
ISCS	.43	1.91	.33	.20	14 - 0
SCIS	.53	2.24	.50	.24	21 - 7
ES	.72				14 - 3

constructed strictly on the basis of content acquisition objectives, this result is not disturbing, especially when we are only interested in the difference between the means of the groups. Incidentally, the reliability coefficents of the control groups are generally sufficiently high. The validity of an instrument is mostly a matter of the correlation between the judgments of a few experts in the field. As Cronbach (1971) remarks, the validity of a test is only its validity for a particular interpretation or purpose. The question in our research would be, "Is this test valid for comparing the mean performance of one group with another?" The leader and instructors of each workshop were consulted as the judges for this type of validity

and in some cases minor changes were made in the test. Although the instructors did not always agree on the validity of an item, they agreed that each Post-Test as a whole was valid.

Data Analysis and Results

The difference between the means of the control and experimental groups was tested using t-tests for independent samples of different numbers. The results of these tests are reported in Table 4 below. The mean score of each curriculum workshop was significantly higher than the mean score of the corresponding control group levels of

TABLE 4

Analysis of Between Group Differences on Posttest Part I

Curricula	Workshops			Control Groups			t
	N	Mean	S.D.	N	Mean	S.D.	
SAPA	44	18.55	3.17	35	14.54	5.04	4.263**
ESS	99	20.21	2.51	43	17.28	4.79	4.735**
ISCS	19	11.63	1.87	57	9.86	2.57	2.733*
SCIS	118	17.17	3.00	43	15.09	3.28	3.779**
ES	131	11.24	3.26	36	8.70	4.00	3.446**

* $p < .005$
** $p < .0005$

significance as high as .0005. This rather great difference between the means suggests that although teachers outside the workshop may have some familiarity with the curriculum, their knowledge is not likely to pass the level of simple facts and concepts about the curriculum. Since we have not given any pre-tests to both the experimental and the control groups, one may argue that the diff-

erence between the means may be due to the original difference between the two groups. This argument does not hold because:

a) the control group was selected randomly from the participants of other workshops of Project ASSIST who are basically as interested in curriculum improvement as the experimental group,

b) other measures obtained at the beginning of the workshops (such as Workshop Participant Perception) do not show any significant difference between one workshop and another.

Therefore, one is led to believe that the workshops have been quite effective as far as the fulfillment of the preceeding objectives are concerned.

One can observe from the data that there are some individuals in each control group who know nothing about that curriculum. The low range score of SAPA, ESS, and ISCS control groups is zero which is lower than the chance score. In the case of SCIS the lowest score was 7 which is close to the chance score ($\sim$ 8) and in the case of ES it was 3 which is lower than the chance score. Here, too, the need for curriculum workshops across the State becomes obvious one more time. They will at least make the teachers familiar with national curricula. It is important for each and every teacher to have a minimum knowledge of major national curricula.

Assessing Affective Outcomes

The measure of attitudes toward the implementation effort, its workshops, and science in general is another important dimension to

be assessed in this project through written measures. Here, however, the type of reporting and data gathering vary from that of the usual objective or simulation type test. The third procedure known as "Workshop Participant's Perceptions Survey" was utilized weekly to monitor participant's attitudes toward the many implementation activities. Instruments which measure attitudes were used frequently during the activity being evaluated in order to determine the attitude relative to the activity. Too often participant perceptions represent only those activities just prior to the assessment, rather than a summative viewpoint of the workshop in general.

Affective Outcomes Objectives Assessed

While many objectives reflecting affective gains and outcomes could be written, the following summarizes the most important of these :

Each curriculum implementation effort will be conducted in a manner which will facilitate positive participant attitudes toward the curriculum. Behaviors which were considered as evidence supporting the attainment of this objective include the following:

(1) Participant and instructor's perceptions of workshop activity were in high agreement.

(2) Workshop instruction facilitated high levels of commitment to use curricular materials.

(3) Instructional procedures were judged by participants and instructors to be relevant to the curriculum.

(4) Participants chose to use curricular materials in their classrooms.

(5) Participants chose to provide evidence of their classroom teaching involving the use of these materials.

The Workshop Participant Perceptions

Procedures for evaluating instruction can generally be classified into one of two categories according to the type of information they utilize. One category is typified by measures of acquisition and retention on instructional content, such as student achievement. Evaluation of this nature follows closely to a cognitive view of instructional effectiveness. A second category utilizes perceptions and impressions to evaluate instruction. These types of data represent voluntary responses to an instructional situation and as such can be considered as affective behavior. Hence, evaluation procedures utilizing this type of information are concerned mostly with instructional affect.

Relationships between instructional effectiveness and affect remain inconclusive. Highly effective instruction is not always highly affective. In spite of this, the literature is filled with instruments attempting to measure instructional affectiveness. Most of these instruments confound the problem with their general lack of a conceptual framework to guide their construction or interpretation. Too often questions used to sample perceptions assume some "ideal instructional model" which may not be congruent to the one utilized

by the instructor. Some questions on these instruments ask for information that the student is not really able to provide. But more important, too many instruments falsely assume that impressions and perceptions are cummulative in a way similar to cognitive learning. Hence, many procedures prescribe only one assessment, usually near the end of the course.

An alternative procedure for assessing instructional affectiveness which attempts to overcome the above constraints evolves from the area of communications theory. While instruction can be described in terms of many factors, communication between instructor and students is a common one. This communication may be formalized through objectives, written materials, evaluation procedures, course work, and assignments. Communication also includes what the students and instructors say and do during instruction. The instructional situation may be a laboratory activity, tutorial teaching, lecturing, group discussion, and others. Whether formal or informal, communication also implies that an interchange takes place not just from the instructor to student but from student to instructor as well. It is not important that the instructor "lectured most of the time". It is important that the instructor knew what the students were doing during the lecture and whether or not their behavior was in tune with the type of expected learning. Communications in any instructional setting will have their maximum influence on student learning only when students attend to them. Their impressions of these communications should be similar to the intentions and impressions of the

instructor. It is the intention of the instrument, "Workshop Participant Perceptions", to assess the nature of the communications network between the students and the instructor as a means to estimate instructional effectiveness.

The instrument, Workshop Participant Perceptions, was developed by the Project to assess instructional affect. It presents a set of items and responses which describe various types of values and activities which may or may not be relevant to the instructional situation. Both the instructor and the students respond to the instrument. The scoring procedure depends upon both the instructor's impressions as well as the students' perceptions. The instructor's impressions are used by the scoring program to assign weights differentially to all responses for each item. These weights are then utilized in calculating a "percent agreement between students and instructor; that is, how closely student perceptions match the impressions of the instructor. While this technique can estimate instructional affect, one set of data can only represent a contemporaneous viewpoint. Repeated assessment is and was necessary to determine general trends and estimates of attitude development.

Data Analysis and Results

While many data points could have been considered, the means and ranges of scores were of particular importance. First, the mean approached 100%, the perceptions of the class more closely matched those of the instructor. Generally, percents around 90% were considered to indicate high agreement. Most of the means were

90% or better; all of the means were better than 80%.

The second statistic, the range of percentages of agreement, displayed a great variance. Ranges indicate more about individual perceptions within the group being considered. When the range was narrow, then most of the group had a similar perception. However, if the range were wide, then many differing perceptions existed. All of the ranges observed had a maximum of 100% with lower limits varying between 40% and 78%. It should be pointed out that wide ranges do not always indicate confusion and diversity of perceptions; consensus may exist even with wide ranges. Generally, group consensus exists when the mean is to either one end of the range; wide ranges with a high or low mean indicate consensus. The data presented in Table 5 all contained relative high means, indicating that group consensus was

TABLE 5

Summary of Means*, Standard Deviations, and Ranges* of Workshop Participant Perceptions**

Curricula	N	Mean	S.D.	Range of Scores
SAPA	44	92.28	5.01	70.3 - 100.0
ESS	100	90.56	4.75	78.0 - 100.0
SCIS	118	91.65	5.11	63.7 - 100.0
ES	131	88.43	5.49	63.7 - 100.0
ISCS	19	90.23	4.75	76.9 - 100.0

* Percent signs omitted.
** Subscales omitted; scores represent an overall score.

rather high also. Probably a few differing perceptions have extended the range much lower than most of the group perceptions. This sus-

picion is confirmed by noting that twice the standard deviation below the mean is still in many cases way above the lower limit.

In general, data supported the objective to foster positive attitudes of the participants toward the curricular materials. They did attempt to present information and generally had similar perceptions of workshop transactions.

Classroom Implementation - Self Reported Data

Part III of the posttest asked participants about their use of materials in the classroom. Generally positive intentions to implement curriculum programs and materials in part or in whole was expressed by 81% of all workshop participants. Furthermore, as reported in Table 6, 91% of all participants expressed a continuing interest in

TABLE 6

Classroom Implementation of Curricular Materials:
Summary of Self-Report Type Data*

Curricula	Implement		Number of Activities					Time/Week (Hours)				
	Yes	No	0	1-4	5-8	9-12	12+	1	1-2	2-3	3-5	5+
SAPA	81	19	11	16	16	7	27	24	41	14	8	0
ESS	79	21	1	44	30	13	8	14	36	22	19	5
SCIS	86	14	7	17	18	13	42	11	36	23	18	4
ES	77	23	12	27	23	11	18	26	21	15	9	19
ISCS	88	12	12	24	18	6	35	9	6	18	12	59

*Percent Signs Omitted.

studying and using the curriculum. These are very encouraging percentages. Although high, they are not surprising since over 50% of all participants considered themselves knowledgeable of the curriculum concepts and techniques and therefore felt competent to implement the

curriculum. However, only 38% of all the participants used on the average one science activity per week from the curriculum program on which they were working, and 78% of all participants spent approximately one-two hours on that activity.

Interpretation

While these data seem to conflict, they are self-reported data and must be interpreted with this in mind. Perhaps it is best to consider these levels more as indications of positive attitude toward a curriculum. Some external factors should be considered. First, district requirements vary as to the amount of time that is to be spent on science instruction. This time varies considerably from one school district to another and teachers can juggle the amount of instruction time from week to week. Hence, the responses should be considered only as rough estimates based on recall of unscheduled activities.

A second factor is that the length of time required to do one unit in one program varies as opposed to one unit in another program. For example, one would suppose that a larger number of ES activities as opposed to SCIS activities could be completed in a given time period, because SCIS is concerned with the type of science investigations that can require more time per unit than many ES activities. The data does not support this later factor. Forty-two percent of SCIS participants taught more than twelve (12) activities per year in 2-3 hours per week. The same percentage of ES participants taught approximately nine(9) activities in 2-3 hours per week. Ob-

viously program differences alone can not predict the amount of activities that workshop participants can teach in a given amount of time. Individual teacher differences have no doubt influenced their productivity. The various degrees of knowledge and understanding of the program in relation to the progress of the workshop would determine the number of activities attempted in the classroom and the amount of time used on these activities.

The inconsistence of self-report type data should be taken seriously. In this set of evaluation data, self-reporting indicates conflict with both performance and knowledge. Self-reporting hence is not a measure of performance or knowledge. The fact that 81% indicated voluntarily that they would implement the curriculum was an indication of a positive attitude toward the curricular materials. Time remains a problem in elementary school programs as it does in enrichment programs such as ES. The self-reported data concerning time and number of activities reflect this concern. All of these findings are, of course, confronted by teacher capability and district tolerance for new innovative approaches to science teaching.

Multidimensional Assessment of Philosophy of Education (MAPE)

The central question of this aspect of the evaluation was the following issue:

> "Do teachers teaching a specific national science curricula exhibit a particular philosophical profile?"

Here it is assumed that a person's philosophy of education represents primarily an affective variable that is reflected in one's pre-

ferences such as teaching strategies on specifying objectives in behavioral terms rather than vague purposes, or presenting content according to academic structures rather than learner sequenced instruction, common in inquiry type activity. Since most science programs prescribe certain modes and styles of instruction, it was expected that differences should occur between the philosophy of a "program", as reflected by selected model teachers.

General Description of the MAPE

The Multidimensional Assessment of Philosophy of Education describes a person's philosophical profile relative to education in a broad sense. This profile is constructed in terms of the following six subscales:

	Subscale:	Reliability:
(1)	Classroom Climate	.85
(2)	Individual Differences	.82
(3)	Teaching Style	.81
(4)	Learning Emphasis	.84
(5)	Procedures and Planning	.81
(6)	Theoretical Base	.84

The test itself presents sixty sets of four items. The instructions direct the examinee to select the single response which he MOST agrees with and then from those remaining, the one he LEAST agrees with. Some items appear in more than one of the sets of items, providing a means to calculate an internal validity coefficient.

Description of the MAPE Subscales

Classroom climate. A high score indicates a perference for unstructured approaches when dealing with classroom activity. Here the teacher would maintain authority more indirectly, involving students in decision making. Low scores indicate strong tendencies toward establishing the teacher as the center of authority in the classroom when decisions would often be predetermined.

Individual differences. A high score indicates an active interest in acknowledging individuals as individuals rather than relying on customary expectations for an individual as a "type" or member of a group. Here teachers would prefer to find out for themselves. A low score suggests a preference for procedures which disregards individual differences in favor of a conventional expectation.

Teaching style. A high score indicates preferences for personal concern for teaching, both in the classroom and professionally. These teachers would tend to be pace-setters rather than followers. Low score indicates impersonal viewpoint where much of the responsibility in decision making is allocated to others, often excluding the actual teacher. Teachers here tend to be more conforming.

Learning emphasis. A high score indicates preferences to maintain a social group learning focus reflecting an emphasis on students' needs as well as continual improvement to meet these needs. A low score indicates a preference for the learning focus to be text-book

oriented where instructors rely primarily on a text-book or other pre-set instructional system. Revision here generally reflects up-dating content in order to meet other external needs.

Procedures and planning. High scores indicate preferences for procedures which adhere to carefully described plans. A systematic approach would be considered here as an effective means to attain desirable objectives. Low scores indicate tendencies to avoid and distrust some of the more conventional planning procedures. Preset procedures are often felt here to be inappropriate.

Theoretical base. High scores indicate an idealistic orientation. Education is viewed as the result of applying theoretical principles that facilitate learning. Heavy responsibility falls to the teacher to see that the child learns a low score indicates a practical or pragmatic focus on the matter of learning. A sound basic approach to teaching is seen as sufficing and fancing theoretical ideas only get in the way.

Identification of "model" teachers. Teachers were identified as a model teacher for a specific curriculum according to these criteria:

(1) They must be currently teaching the target curriculum.

(2) They must have attended some special in-service or summer workshop relative to teaching the target curriculum.

(3) They must be teaching the curriculum in the style intended by the curriculum developers. Evidence for this was on the basis of audio-tape recorded samples of their teaching.

While a total of 600 teachers were considered as the sample initially, only 185 met all three of the above criteria. Table 7 indicates the number of teachers for each curriculum that were finally included.

Interpreting MAPE Scores

Scores for each of the above six subscales are reported as standard T-scores with a mean of 50 and a standard deviation of 10. However, the mid-point for all subscales is also 50. Scores are based on a referent group of over 500 teachers and administrators. The set of six scores can be thought of to describe a profile of a person's philosophy of education. Tendencies and preferences should be considered stronger the further a score moves above or below the middle scale point of 50.

Results

The results of the MAPE are reported in Table 7 on the following page. Using these scores, a multiple discriminate analysis was performed in order to determine which of the subscales best discriminates between each of the curricular philosophical profiles. Table 8 reports the discriminate scores.

Using a level of significance of .05, three subscales were included in the linear discriminate equation. These subscales were in order of inclusion learning emphasis (subscale 4), Procedures and Planning (subscale 5) and Teaching Style (subscale 3). The remaining subscales failed to pass the necessary level of significance. The discriminant scores reported in Table 7 may be used to match educators

to a curriculum of best fit. Using their MAPE subscores 4, 5, and 3 and the discriminant scores and coefficients reported in Table 7, a personal discriminant coefficient would be calculated for each curriculum. The curriculum with the lowest discriminant coefficient would be the one of the best match for that person.

Analysis

It is interesting to note that the major differences between the curricula implemented and described by their learning emphasis, procedures and planning, and teaching style. Curricula such as ESS and TPE strive to maintain a social-group focus, rather than text-book. This finding is congruent with the curricular material. However, IPS and SAPA would seem to have a heavy reliance on a predetermined instructional system, a finding also congruent with the materials. Most of the scores reflect an adherence to described plans, no doubt a finite of the sample itself. However, teaching styles offer some additonal insights from ESS and SAPA was extremely high (relative to the group) indicates teachers held a preference for personal concern for teaching activities. None of the curricula disclosed a "low score.

Analysis of this type may have utility in identifying the extent to which teacher preferences match those who are model teachers of the curriculum. Those who have a mismatch could potentially benefit most from a more thorough in-service program.

TABLE 7
MAPE Philosophical Profiles for Seven National Science Curricula

Curriculum Population:	N		(1) Classroom Climate	(2) Individual Differences	(3) Teaching Style	(4) Learning Emphasis	(5) Procedures & Planning	(6) Theoretical Base
ESS	50	M=	49.73	55.02	61.84	51.64	55.03	57.91
		SD=	8.22	9.34	9.99	10.66	11.32	9.64
ES	50	M=	47.26	52.45	58.90	50.02	55.39	58.83
		SD=	10.95	10.19	10.94	11.18	10.56	9.58
SCIS	20	M=	48.60	53.62	59.40	49.44	53.47	59.41
		SD=	10.69	8.83	11.33	10.28	9.47	10.16
SAPA	20	M=	47.42	53.45	60.56	45.52	51.63	59.85
		SD=	10.08	8.39	11.53	12.66	8.04	9.20
ISCS	20	M=	46.54	51.12	57.84	49.32	54.90	59.95
		SD=	9.22	10.85	10.62	10.08	10.17	7.24
TPC(ECCP)	15	M=	47.62	51.44	55.38	52.96	55.68	55.81
		SD=	8.67	9.54	10.72	11.02	7.75	11.11
IPS	10	M=	41.57	50.43	56.16	44.94	53.53	54.74
		SD=	10.70	9.81	16.11	13.64	10.87	10.67

TABLE 8
Discriminant Scores* with Subscales 3, 4, & 5 for Curriculum Populations

Curriculum	Subscale 4 (b_4)	Subscale 5 (b_5)	Subscale 3 (b_3)	Constant (c)
ESS	0.317256	0.532191	0.485553	-39.828583
ES	0.305839	0.536306	0.464254	-38.122528
SCIS	0.303788	0.507433	0.467426	-36.651016
SAPA	0.270282	0.504051	0.480227	-35.651993
ISCS	0.301399	0.531526	0.455913	-37.151550
TPE(ECCP)	0.333900	0.534024	0.432685	-37.636826
IPS	0.267241	0.521838	0.445637	-34.430832

*Linear Discriminant Equation: $D = b_4 x_4 + b_5 x_5 + b_3 x_3 + c$

Assessing Participants' Technical Skills

A question basic to the success of this project is concerned with the manner in which participants implement curricular materials in their classrooms. In many cases, curricula specify teaching situations and techniques not commonly used in more traditional types of instruction. Hence, the growth of technical skills on the part of the participant is an important assessment. Since the period of the project extended through one calendar year, the assessment procedures needed to provide a vehicle for instructors to monitor participant classroom efforts more frequently than would be possible if the monitoring were limited to visitations to participant classrooms.

Three approaches to assessing technical skills were utilized. One method consisted of a discrimination task in a pretest-posttest assessment. Another analyzed audio-taped examples of classroom teaching. A third relied on self-report techniques. Each of these methods utilized a different, but appropriate, design. Each of these methods will be dealt with separately.

Objectives Assessed

The following objectives were the target of the technical skills assessment.

A. When teaching lessons from a specified curriculum project (i.e., ESS, SAPA, SCIS), the committed teacher will utilize those verbal and non-verbal teacher behaviors most relevant to the goals of the curriculum.

Behaviors which are evident in supporting this objective included:

(1) The teacher used appropriate skills in teaching lessons from the curricula.

(2) The teacher planned and carried out lessons using specific teaching skills.

B. As a result of the technical training sequence, the committed teacher will exhibit competency by adequately performing (in terms of frequency and quality) the teacher behaviors identified as relevant to the curriculum.

(1) Teacher can recognize and identify specific inquiry related teaching skills.

(2) Teacher will provide audio-tapes of classroom teaching demonstrating their capability to use selected teacher skills in implementing curricular materials.

Pretest-Posttest Assessment

An instrument was developed which presented samples of teacher questions requiring the participant to identify the teaching skill represented by the question. For example, a sample test item could have been, "What color is the block?" Correct responses would identify the question as one that elicits observation. It is important to note that tests of this type depend upon presenting the novel questions, new to the participant, as a means of determining whether or not the participant has acquired some comprehension of the nature of the teaching skill. As Bandura (1971) points out,

acquisition of a model behavior for performance at another time, such as in classrooms, without the model of the teaching skill present, requires that observer learn the behavior and remember it in a symbolic form according to the characteristics which identify that skill from other skills. The instrument utilized to assess this type of learning is congruent to both Bandura's writing and prior research in the area (Koran, 1969; Wilson and Koran, in press).

Design. A random sample of participants were selected during May, 1974, and administered the instrument, "Teaching Skills Inventory--Part III". The same people, with some attrition, took the instrument again a year later. A t-test for correlated data was used to evaluate observed differences.

Results. Significant differences as reported in Table 9 were found between pre-test scores and posttest scores at the .01 level. Pretest-posttest designs can be criticized for "tuning the sample" and hence accounting for observed increases in scores. However, the differences recorded here and the other data assessing teaching skill, all indicate that the increase in teacher skills was ot due to just "pretest influence".

TABLE 9
Pretest-Posttest Analysis for
Teaching Skills Inventory -- Part III

Test	Mean	Standard Deviation	N	t
Pretest	13.37	3.49	46	
Posttest	16.02	3.12	46	
Difference	2.65	4.56	46	+3.9468*

*p < .01

Audio-Tape Analysis

All workshops required participating teachers to record portions various attempts to use curricular materials in their classrooms. These audio-tapes were then analyzed by the participant and the workshop instructor as a means of monitoring classroom teaching as well as facilitating the acquisition of essential teaching skills. These audio-tapes were then collected by the assessment and evaluation staff; analysis of these tapes provided a still different source of data. The audio-tape permitted an assessment of teaching style and skills as they occurred in nearly normal classrooms.

Design. The repeated sampling of classroom teaching falls into the quasi-experimental category of research designs. The major objective of these data was to examine the trends relative to teachers using the various categories of teaching skills.

Procedures. Special instructional materials were prepared describing to both instructor and participants the mechanics for making and analyzing audio-tapes. Models of each of eleven skills were constructed. These skills correspond to those identified as relevant to inquiry-type activities. After teachers made an audio-tape, they self-rated their own performance. Then they discussed their recorded performance and self-rating with their instructor. When both participant and instructor were finished with the audio-tape, it was sent to the assessment and evaluation staff.

Hundreds of audio-tapes were received throughout the school year. Two random samples of these tapes were made, one in the fall and one

in the spring. Five minute samples of the recorded teaching were randomly copied from each selected tape. These five minute samples were then rated by two trained rators. Estimates of relibability were calculated using analysis of variance techniques. In most cases, reliability was acceptable (> .65).

Results. Four skills, as reported in Table 10, were found to occur in relatively high frequency (greater than once in five minutes); these included reinforcement, probing, eliciting observations,

TABLE 10
Audio-Tape Analysis for Inquiry Teacher Skills

Teacher Skills	Random Sample 1*			Random Sample 2**		
	M	SD	Rel.	M	SD	Rel.
Reinforcement	1.017	2.314	.91	1.171	2.975	.85
Silence	0.300	0.949	.57	0.114	0.577	.08
Probing Questions	2.300	4.276	.97	2.743	5.504	.83
Procedural Questions	1.517	3.315	.85	0.537	1.977	.86
Elicits:						
Observations	2.350	4.584	.93	3.671	7.035	.83
Classifications	1.100	2.306	.87	1.043	3.027	.71
Inferences	0.783	1.800	.91	0.729	2.111	.90
Predictions	0.617	1.672	.85	0.886	2.534	.80
"Measurements"	0.317	1.331	.95	0.686	2.157	.94
Simplifications	0.400	1.189	.91	0.057	0.333	.13
Relationships	0.083	0.401	.56	0.080	0.707	.04

*N=40 **N=35

and classifications. All the remaining teaching skills occurred less than once in five minutes. This is not unexpected since many of these behaviors represent higher-order questions which would by their nature be included in classroom dialogue less frequently.

Randomly sampling of the recorded teaching activity was an essential procedure since many teaching skills are performed differentially depending upon when in the lesson the recording is made. In other words, teachers tend to "probe" more in the beginning and end of a lesson rather than at the beginning. One goal of this analysis was to obtain a general description of how inquiry teaching skills were used in classroom situations; differences due to "when" the skills were utilized in teaching needed to be minimized. Certainly the "when" is an important consideration, but sequence of skill use was not a central question. By randomly selecting segments of teaching, all behaviors had a more nearly equal opportunity of occurring, unless they simply occurred more often in classrooms. Hence, the profiles found for Random Sample 1 and Random Sample 2 can be said to represent science teaching in general as it occurred in participating teacher's classrooms.

Self-Report Assessment

As part of the post-test part III, which was mailed to all participants as described in the previous section, "Assessing Participant Achievement", participants were asked to rate how well they had acquired and performed skills in various categories of teacher behaviors. The rating scale was a five point scale as reported below:

(1) Behavior unknown to me;
(2) Behavior familiar to me;
(3) Behavior exhibited before workshop experience;
(4) I tried to acquire the behavior but didn't;
(5) I acquired the behavior as a result of the workshop.

For the purposes of this analysis, Response 1 and 4 were combined. The results of this analysis revealed some interesting patterns, as presented in Table 11. Generally those categories of responses can

TABLE 11
Inquiry Teaching Skills:
Summary ofSelf-Reported Competencies

Teacher Skill	Level of Performance	Level of Competency: Skill Exhibited Before	Know Skill Now	Acquired Skill Now	Still Don't Know
Reinforcement Probing Observing Classifying	High Frequency	50%	40%	10%	2%
Measuring Inferring Predicting Simplifying Stating Relationships	Low Frequency	25%	30%	34%	10%

be divided into two groups. The first group contain those behaviors that teachers feel they already use in the classrooms before any experience with the project activities. These behaviors include reinforcement, probing, eliciting observation, classification, and measurement. Generally, these behaviors were found in audio-tape analysis to be those which occur most frequently. The second group included inquiry teaching skills where a large percentage of participants acquired through the workshop activities. These included eliciting inferences, predictions, simplications, statements of relationships. These behaviors are also those which tend to occur

less frequently in samples of classroom teaching. Few teachers reported that they were still unfamiliar with the inquiry behaviors after the workshop.

Summative Viewpoints

The descriptive aspects of this evaluation presented a variety of data illustrating the complex nature of the Project ASSIST implementation workshops. These diverse activities were to all contribute to improving classroom science teaching. In summarizing whether or not the activities have in fact made an impact on that classroom instruction, two possible views can be expressed -- namely that of an advocate and an adversary. As Evans (1974, p. 8) points out,

> "Since virtually all education programs have their committed advocates and strong detractors, and are in this sense inherently controversial, it is inevitable that evaluations of them will also be controversial. Any evaluation that finds a program successful will be attacked by the program's detractors, and any evaluation that finds the program unsuccessful will be denounced by its advocates."

The challenge accepted in this report is to pull together an adversary and an advocate view. This does not mean that the authors here are "for" or "against" the project, but rather they are interested in identifying its strengths and weaknesses.

Advocates View

A propensity of data was presented describing in many ways the effectiveness of the in-service workshop. Teachers gained in knowledge, skills and attitude. The skills acquired were congruent with the

nature of the curriculum. They did, in fact, attain the capability to utilize the curricula in meaningful ways. Evidence supporting this was not only direct evidence via recordings of classroom activity, but also includes samples of student work and reports from teachers of their efforts.

Assessment of Curriculum Relevant Knowledge. The information that teachers need to know in order to use a curriculum extends beyond the facts and concepts contained within the curriculum. They also need to recognize and understand some of the assumptions and instructional design elements basic to the curriculum; this knowledge should guide their implementation. The assessments have revealed that most workshops have been quite effective in helping teachers acquire curriculum relevant knowledge.

Affective Outcomes. A wide variety of information was sampled relative to attitudes and affective type responses. Generally, the data supported the objective to foster positive attitudes of the participants toward using and implementing the curricular materials. Most have expressed strong intentions to continue their efforts to use the materials even after the workshop was concluded. Certainly the reaction of experienced teachers must be respected, especially when such a high percentage repeated that they would continue to utilize these materials in the future.

Effectiveness of Teacher Training Materials. A major objective of the project was to facilitate the acquisition and use of inquiry related teacher skills. The data presented generally supported the use of specially designed materials and procedures. Further contrasts of the data, as presented in Table 12,reveals even a stronger relationship between participant involvement with these materials, instructional feedback, and the resulting acquisition of selected teacher skills. The most significant gains in teacher skill, (using

Table 12

Level of Involvement vs. Teaching Skill Competency:

Curricula	N	Audio-Tapes*	Classroom Teaching*	Helpful Feedback**	Posttest Recognition of Teaching Skills Mean	S.D.	d/F test
ISCS	20	60%	72%	83%	26.35	6.54	d = 1.00
SAPA	45	21%	22%	88%	25.04	8.86	d = 0.48
ESS	100	89%	83%	87%	27.99	7.60	d = 2.84*
ES	132	14%	25%	85%	24.31	8.35	d = 0.07
SCIS	119	81%	92%	85%	28.15	8.02	d = 2.85**
Control	51	(No training)			24.27	8.42	
Total	467						F = 4.809 ****

* Based on materials completed and returned by June 1, 1975
** Based on Workshop Participants Perceptions (WPP) Assessment.
*** ($p < .01$)
**** ($p < .001$)

Posttest Part II scores), occurred in curricula where high involvement occurred with helpful feedback. These elements represent practice opportunity and feedback, both previously found important in effective teaching and learning situations. These materials and procedures

should be considered as an effective, on-the-job, retraining procedure to help teachers up-date their teaching skills.

In summary, the advocate asserts that the in-service workshops more than met their objective. Without a doubt, the workshops also presented new ideas and stimulated changes. In this aspect they went beyond their objective and provided a valuable service throughout the state of Iowa. In short, the ASSIST effort promoted quality classroom science teaching.

Advisaries View

Projects of this nature seem intent on capturing large numbers of persons and organizing them into some quick, but impressive activity. Certainly the members here are impressive. But fundamental to this is whether or not the activity suits their needs or does the activity create their needs. If needs are very fulfilled by activity, then impact should be long range; it should exist beyond the year-long workshop. No evidence describing "after workshop" impact was gathered; improved instruction may or may not now exist.

Improving instruction is an all-to-often used expression for describing what was done to teachers rather than by teachers. The real measure of improved teaching must be in terms of the learner. Was what the teacher did appropriate from a learner's perspective? Data included in this report indicates that what teachers did was appropriate with respect to the curricula. However, the question seems to be, are the selected curricula appropriate? Do the curricula selected by the school and teacher fill their needs or

the learner's needs? This is a basic question; one that should be answered first. While it may also be one that goes beyond the scope of the project, from this viewpoint here it would seem to be a reasonable responsibility.

In summary, the adversary contends effective activities described in this report are in many cases fulfilling "empty" and "inappropriate" needs. Workshops were sold as a means to create activity rather than find and provide assistance with needed classroom instructional problems. No evidence suggests otherwise.

Conclusion

Long range effects of implementation effort such as Project ASSIST remain yet unknown. While this implementation effort has been successful even beyond a simple head count, its effects through time need to be examined. In addition, effectiveness needs to be evaluated in still another way (Evans, 1974, p. 8):

> "Once, evaluations move beyond the primitive efforts to conduct site visits or simply collect data on a before-after basis, the evaluator and the design he employs must confront the fundamental problem of providing some estimate of what would have happened in the absence of the program he is evaluating."

In this regard a careful examiner would do well to take a wide view and include not only the obvious effects such as whether or not participants maintained or improved their use of materials in the classroom, teaching skills, and new curriculum development and implementation, but also secondary effects such as increased university and local school associations relative to various educational

concerns, closer association between universities and area educational agencies and state departments of instruction, and perhaps even increases in enrollments in various advanced degree or certification programs. Evaluation of this type is the responsibility not only of Project ASSIST, but also of the supporting universities, national agencies, and participating schools. Perhaps in this way, future decisions relative to educational change will prove to be even more effective.

References

Bandura, A. Psychological Modeling. Chicago: Aldine-Athertone, 1971.

Cronbach, L. J. Test validation. In R. L. Thorndike (Ed.), Educational Measurement, Washington,D. C.: American Council on Education, 1971.

Evans, J. W. Evaluating educational programs - are we getting anywhere? Educational Researcher (September, 1974).

Koran, J. J., Jr. Supervision, an attempt to modify behavior, Educational Leadership, 26 (8), 1969

Wilson, J. T. and Koran, J. J., Jr. Teacher Corps Preservice: A Study of Change in Science Teacher Behavior. Journal of Research in Science Teaching, In press.